



**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE BIOLOGIA MOLECULAR
CURSO DE BACHARELADO EM CIÊNCIAS BIOLÓGICAS**

MATHEUS AUGUSTUS TEIXEIRA SILVA

**UTILIZAÇÃO DE AMINOÁCIDOS NAS PROTEÍNAS PRIMITIVAS E A
ORIGEM DO CÓDIGO GENÉTICO**

**JOÃO PESSOA - PB
2018**

MATHEUS AUGUSTUS TEIXEIRA SILVA

**UTILIZAÇÃO DE AMINOÁCIDOS NAS PROTEÍNAS PRIMITIVAS E A
ORIGEM DO CÓDIGO GÉNÉTICO**

Monografia apresentada ao curso de ciências biológicas (Trabalho Acadêmico de Conclusão de Curso) como requisito parcial para a obtenção do grau de bacharelado em ciências biológicas da Universidade Federal da Paraíba.

Orientador: Prof Dr: Sávio Torres de Farias.

**JOÃO PESSOA - PB
2018**

S586u Silva, Matheus Augustus t.

Utilização de aminoácidos nas proteínas primitivas e a origem do código genético / Matheus Augustus t Silva. - João Pessoa, 2018.
48 f. : il.

Orientação: Sávio Torres de Farias.
Monografia (Graduação) - UFPB/CCEN.

1. RNA transportador. 2. proteínas. 3. bacteria. 4. Archeae. 5. Eukarya. I. de Farias, Sávio Torres. II. Título.

UFPB/CCEN



UNIVERSIDADE FEDERAL DA PARAÍBA

Centro de Ciências Exatas e da Natureza

Coordenação do Curso de Ciências Biológicas

Telefone: (083) 3216.7439, Fax (083) 3216.7464.

CEP 58059-900 - João Pessoa, PB, Brasil. e-mail: cccb@dse.ufpb.br


**Ata da Apresentação e Defesa de Trabalho
Acadêmico de Conclusão de Curso da
Estudante Matheus Augustus Teixeira Silva**

Aos **quatorze** dias do mês de **junho** de dois mil e dezoito, na sala da Pós-graduação do **DBM** da Universidade Federal da Paraíba, Campus I, João Pessoa-PB, reuniu-se, em caráter de solenidade pública, às **treze** horas, a Banca Examinadora do Trabalho Acadêmico de Conclusão de Curso do estudante **Matheus Augustus Teixeira Silva**, composta pelos seguintes professores doutores: **Prof. Dr. Sávio Torres de Farias/ Orientador e Presidente da Banca Examinadora, Profa. Dra. Louisa Maria de Andrade e Sousa/ Examinador e Prof. Dr. Augusto Cezar Vasconcelos de Freitas Junior Examinador**. Compareceram à solenidade, além do estudante e dos três membros da Banca Examinadora, alunos e professores do Curso de Graduação em Ciências Biológicas da Universidade Federal da Paraíba. Dando início à sessão, o Senhor Coordenador do Curso **Prof. Dr. Tarcísio Alves Cordeiro** fez a abertura da sessão, apresentando o estudante. Em seguida fez a apresentação da Banca Examinadora e passou a presidência da sessão para **Prof. Dr. Sávio Torres de Farias** que, concomitantemente, assumiu a posição de orientador e presidente da sessão que, após declarar o objeto da solenidade, concedeu a palavra a estudante, candidato ao Grau de **Bacharel** em Ciências Biológicas, para que dissertasse, oral e sucintamente, a respeito do trabalho de título **“Utilização de Aminoácidos nas Proteínas Primitivas e a Origem do Código Genético”** Passando então a discorrer sobre o referido tema, dentro do prazo legal, o estudante foi a seguir arguido pelos examinadores na forma regimental. Em seguida, passou a Comissão, em caráter secreto, a proceder à avaliação e julgamento do trabalho, concluindo por atribuir-lhe as seguintes notas: **Prof. Dr. Sávio Torres de Farias 9,0, Profa. Dra. Louisa Maria de Andrade e Sousa 9,2 e Prof. Dr. Augusto Cezar Vasconcelos de Freitas Junior 10**. Com média final **9,4**. Perante a aprovação, declarou-se o estudante legalmente habilitado a receber o Grau de **Bacharel** em Ciências Biológicas. Nada mais havendo a tratar eu, **Sávio Torres de Farias**, como Presidente, lavro a presente Ata que, lida e aprovada, assino juntamente com os demais membros da Banca Examinadora.

João Pessoa, 14 de junho de 2018


Prof. Dr. Sávio Torres de Farias

Orientador


Profa. Dra. Louisa Maria de Andrade e Sousa
Titular


Prof. Dr. Augusto Cezar Vasconcelos de Freitas Junior
Titular

Dedico este trabalho aos meus pais Uthant Sartunino e Rogeria Teixeira e a minha avó Abinada de Caldas.

AGRADECIMENTOS

A Savio Torres de Farias pela orientação.

A todos os membros do laboratório de Genética Evolutiva Paulo Leminski.

As minhas tias Erica de caldas e Sueine de caldas.

Em memória aos meus avós Manoel Edmilson e Maria Teixeira.

A minha irmã Nathalia Caldas.

A toda minha família Caldas e Teixeira.

A minha namorada Thays Xavier.

A Universidade Federal da Paraíba e ao Departamento de Biologia Molecular.

A banca por ter aceitado participar da avaliação desta pesquisa.

A mim por ter concluído esta pesquisa.

“Nada na biologia faz sentido exceto a luz da evolução ”

(Theodosius Dobzhansky)

RESUMO

Entender como surgiu o código genético e as proteínas sintetizadas por ele é de extrema importância para compreender a evolução da primeira célula e consequentemente do primeiro organismo, o LUCA. As teorias sobre a evolução do código genético argumentar principalmente sobre as primeiras ligações entre as moléculas de tRNA e os aminoácidos. Entre as teorias estão a de Francis Crick sugerindo que o código teria sido congelado, a de Carl Woese que propõem a teoria estereoquímica, o modelo de minimização de erros, o modelo do Wong sobre a coevolução, o modelo de código genético do Hartman, e o modelo autoreferente de Romeu Guimarães. Nesse trabalho nos procuramos analisar quais foram os primeiros aminoácidos a entrar no código genético e compor as proteínas primitivas, baseados em análises estatística (médias e variâncias) dos aminoácidos das proteínas traduzidas de tRNA ancestrais retirados FARIAS et al (2016) e dos genomas dos domínios da vida (*Bacteria*, *Archaea* e *Eukarya*), retirados do programa online codon usage database (www.kasusa.or.jp/codon). A partir dos resultados, nos sugerimos que a alta média do aminoácido arginina não prebiótico nas proteínas derivadas da tradução de tRNA ancestrais deve-se a alta média de serina (prebiótico) também alta nas proteínas derivadas da tradução de tRNA ancestrais. E que a alta média do aminoácido lisina (não prebiótico) nos domínios da vida (*Bacteria*, *Archaea* e *Eukarya*) se deve ao um mecanismo de minimização de erro onde uma mutação na segunda base do códon da arginina resulta em lisina. As altas médias para aminoácidos do tipo R apolares alifáticos fortalece o modelo estereoquímico. Os altos níveis de alanina nos domínios da vida devem-se as substituições que ocorreram no processo evolutivo do aminoácido glicina (médias altas nas proteínas derivadas da tradução dos tRNA) por alanina. Os aminoácidos que apresentaram maiores índices de variância foram de grande importância para a entrada de aminoácidos hidrofóbicos e da fase 1 no código genético.

Palavras-chave: RNA transportador, proteínas, *bacteria* *Archeae*, *Eukarya*

ABSTRACT

Understand how did the genetic code and how proteins synthesized by it is of extreme importance for understanding the evolution of the first cell and consequently the first organism, the LUCA. As theories about the evolution of the genetic code mainly argue about the first links between as molecules of the tRNA, and the amino acids. Among such theories are one of Francis Crick suggesting that the code would have been frozen, Carl Woese who propose a estereoquímica theory, the model minimization of errors, the model of the Wong on a coevolução, the model of the genetic code of the Hartman, and the Romeo Guimarães autoreferente model. In this work we what to were the first amino acids into the genetic code and compose as primitive proteins, based on statistical analysis (means and variances) of amino acids of the protein translated from tRNA ancestry withdrawn FARIAS et al (2016) and the genomes of life domains (Bacteria, Archaea and Eukarya), removed from the database program (use of CODON online www.kasusa.or.jp/CODON). The journey of the results, we suggest that the high average of the amino acid arginine does not prebiotics in proteins derived from the translation of tRNA ancestry should become a high average of serine (prebiotics) too high in protein derived from translation of tRNA ancestry . And that a high average of the amino acid lysine (not prebiotic) in the fields of life (Bacteria, Archaea and Eukarya) is a mechanism for minimizing error where a mutation in the second base of the codon of arginine results in lysine. As high averages for amino acids, aliphatic type R apolar strengthens the estereoquímico model. The high levels of alanine in the realms of life must become as replacements that have occurred in the evolutionary process of the amino acid glycine (medium-high in protein derived from the translation of the tRNA) by alanine. The amino acids that higher rates of variance was were of great importance for the entry of hydrophobic amino acids and the phase 1 in the genetic code.

Keywords: conveyor RNA, proteins, *bacteria*, *Archeae*, *eukarya*.

Sumário

1 INTRODUÇÃO	12
1.1 A Química Prebiótica	12
1.2 Estruturas do Código Genético	14
1.2.1 Características do Código Genético	14
1.2.2 Moléculas Responsáveis pela Tradução do Código Genético	16
1.3 Evolução do Código Genético	21
2 OBJETIVOS	26
2.1 Objetivo Geral	26
2.2 Objetivos Específicos	26
3 METODOLOGIA	27
3.1 Contagem dos Aminoácidos das Proteínas Traduzidas de tRNA Ancestrais	27
3.2 Genomas dos Organismos Domínios da Vida (<i>Bacteria</i> , <i>Archaea</i> e <i>Eukarya</i>)	28
3.3 Médias dos tRNA Ancestrais e Domínios da Vida	29
3.4: Médias Gerais por Classes Químicas de Aminoácidos	29
3.5 Análises das Variâncias	30
4 RESULTADOS	31
5 DISCUSSÃO	35
6 CONCLUSÃO	42
REFERENCIAS	43

LISTA DE ILUSTRAÇÕES

PÁGINA

Figura 1 - Aparelho de descargas elétricas empregadas por Miller e Urey 12

Figura 2 - Relação produto-percussor dos aminoácidos segundo Wong 22

Figura 3 - Ácido aspártico e ácido glutâmico: Primeiros aminoácidos a compor o código genético primitivo segundo Wong..... 23

Figura 4 - Modelo autoreferente..... 25

Figura 5 - Proteínas advindas da tradução de tRNA ancestrais.....27

Gráfico 1 - Médias dos aminoácidos advindos de proteínas traduzidas de tRNA ancestrais e dos três domínios da vida. 31

Gráfico 2 - Médias aritméticas dos aminoácidos por classe química: tRNA ancestrais e domínios da vida 33

Gráfico 3 - Variâncias: tRNA ancestrais e domínios da vida. 34

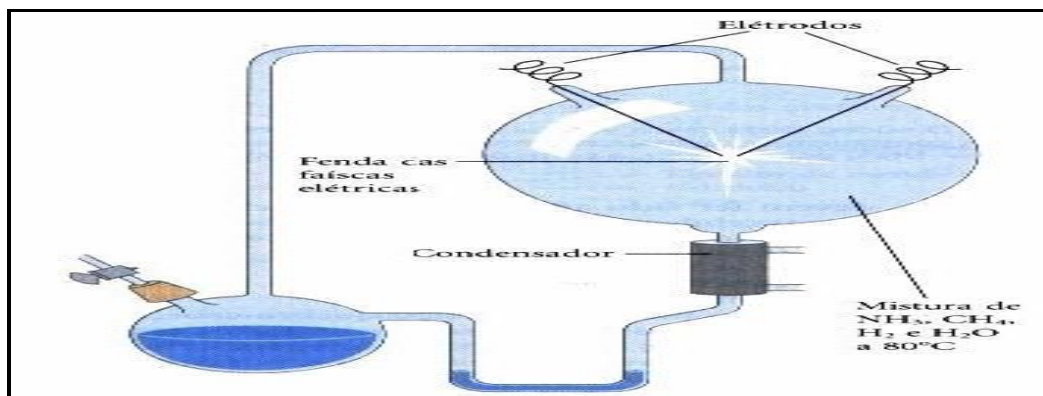
1 INTRODUÇÃO

1.1 A Química Prebiótica

Os primeiros trabalhos sobre a origem da vida em condições prebióticas procederam com o bioquímico Russo Aleksander Ivanovich Oparin e o geneticista britânico John B. Sanderson Haldane. Oparin verificou que a união de moléculas inorgânicas pode originar moléculas orgânicas como proteínas em um ambiente aquoso, e que as moléculas de proteínas poderiam se unir formando estruturas chamadas coacervados. Já Haldane propôs o primeiro modelo de oceanos ou lagoas da atmosfera primitiva chamada de “sopa prebiótica” onde moléculas orgânicas como proteínas apresentavam capacidade de se auto-organizar (KAUFFMAN, 2011).

Baseado nessas ideias, Stanley. L. Miller e Harold. C. Urey realizaram um experimento simulando a atmosfera primitiva da terra, em que misturas de gases CH_4 , NH_3 , H_2 aquecidos com H_2O , sob a ação de descargas elétricas assemelhando-se a relâmpagos, produziram após algum tempo, ácidos graxos, aminoácidos, açúcares, ácido fórmico e ácido cianídrico, Figura 1 (LAZCANO; BADA, 2003).

Figura 1: Aparelho de descargas elétricas empregadas por Miller e Urey



Fonte: Lehninger, 2002

Em experimentos similares ao de Miller e Urey, contudo, usando-se técnicas de análise química mais precisas, evidenciou-se a presença de uma molécula chamada formamida, base para a produção de aminoácidos, bases de ácidos nucleicos, ácidos carboxílicos, uridina e açúcares (PINO et al., 2015). Já em outros experimentos também foi possível ter a formação espontânea de formamida a partir de amônia e dióxido de carbono.

Sob a ação de descargas elétricas a molécula de formamida era capaz de produzir o aminoácido glicina, (FERUS et al., 2017). Em experimentos prebióticos que tentam obter a formação de nucleobases, foi observado que quando simulando a composição da atmosfera primitiva associada a carbonato de cálcio, pirofosfato inorgânico e formamida, submetidos a ação de radiação ultravioleta (UV), era possível obter nucleobases que incluíam purinas, adenina e guanina (BARKS et al., 2010).

Pesquisas realizadas por Chyba et al., (1990); Civiš et al., (2004) analisam em que condições a atmosfera primitiva possibilitou a origem da vida, se redutora ou neutra. Posteriormente, foi documentado que poderia ocorrer a presença de uma atmosfera neutra, através da descoberta do estado de oxidação de rochas ígneas, (TRAIL; WATSON; TAILBY, 2011). Entretanto, outras análises dessas rochas ígneas mostraram que a crosta do hadeano (período geológico que muitos cientistas consideram ter condições propícias para a origem da vida) era redutora (YANG; GAILLARD; SCAILLET, 2014).

Considerando que na atmosfera primitiva os elementos para a formação de DNA, RNA e proteínas poderiam estar presentes, um dos maiores desafios da ciência é saber que macromolécula se estabilizou primeiro no ambiente? Muitas especulações apontam para DNA, RNA ou as proteínas.

Contudo, ao analisar as características químicas de cada uma dessas moléculas, existem fortes evidências de que o RNA foi o primeiro material genético a ser formado, face aos caracteres que constituem essa molécula, quais sejam: maior versatilidade química, pois ela pode ser uma molécula informacional como também possuir atividade catalítica, diferentemente do DNA que só possui atividade informacional e das proteínas que só possuem atividade catalítica e estrutural (DWORKIN; LAZCANO; MILLER, 2003).

Em 1982 foi descoberto moléculas de RNA com capacidade catalítica, as ribozimas, com capacidade de catalisar mudanças em outras moléculas de RNA clivando sequências nucleotídicas, ligar outros dois RNA ou também ajudar a catalisar um processo de polimerização de outro RNA, portanto novas variantes de RNA vão surgindo ou seja as moléculas de RNA podem passar por um processo de seleção natural, fortalecendo assim a hipótese do mundo de RNA (MARTIN; UNRAU; MÜLLER, 2015; MCKENNEY; ALFONZO, 2016).

1.2 Estrutura do Código Genético

1.2.1 Características do Código Genético

A hipótese “um gene uma enzima” proposta por Beadle e Tatum (1940), preconizou a ideia de que era função do gene produzir uma proteína, esta hipótese unificou a bioquímica com a genética. Posteriormente com a descoberta da estrutura tridimensional do DNA, percebeu-se que havia alguma relação entre as sequências dos nucleotídeos e os aminoácidos das proteínas, fato que suscitou questionamentos como: As sequências do DNA determinam as sequências de aminoácidos nas proteínas? A partir dessa pergunta a analogia de um código veio de imediato, os nucleotídeos são as letras e uma combinação dessas letras geram os aminoácidos.

Entretanto outras perguntas foram sendo realizadas, por exemplo: como o código é lido? Ele é superposto ou não, quantas letras no RNA mensageiro (mRNA) constituem uma palavra ou códon e que códon ou códons representam cada aminoácido? Em 1961, através de experimentos genéticos onde foram induzidas mutações em bactérias foi determinado que o código genético era não superposto, pois foi visto que apenas um aminoácido por vez muda de uma região da proteína, já em um código superposto a mudança de uma única base modificará até três aminoácidos em posições adjacentes.

O código genético é composto pelas bases **A, U, G, C** que por analogia constituem as letras e as palavras constituem os aminoácidos, devido a isso se as palavras que originam os aminoácidos fossem constituídas por uma única letra seriam possíveis apenas 4 palavras no código genético. Sabe-se hoje que existem 64 códons possíveis e, portanto o código é redundante (WATSON et al., 2015).

Em 1961, Francis Crick e Sidney Brenner realizaram um experimento com bacteriófago T4, e demonstraram que o código genético é formado por três nucleotídeos (CRICK; BRENNER, 1961). Depois dessa descoberta os pesquisadores queriam saber por que havia tantos códons para poucos aminoácidos? Através desse trabalho Crick sugeriu que o código era redundante, ou seja, alguns aminoácidos são especificados por mais de uma trinca, e depois de ter argumentado isso a redundância foi comprovada através de experimentos bioquímicos (SHABALINA; OGURTSOV; SPIRIDONOV, 2006).

Com a descoberta da redundância, o próximo passo foi saber quais trincas especificam quais aminoácidos, então no mesmo ano, Marshall Nirenberg e Heinrich Matthaei sintetizaram um mRNA mensageiro contendo apenas nucleotídeos de uracila, em seguida misturaram com uma maquinaria de síntese de proteínas de *E.coli*, consequentemente observaram a síntese de uma proteína onde a sequência primária dela era composta por uma fileira de polifenilalanina, assim sendo foi comprovado que UUU especifica a fenilalanina (NIRENBERG; MATTHAEI, 1961). A posteriori destes experimentos foram descobertas novas trincas dos aminoácidos restantes.

Com a decifração do código genético a pergunta a ser feita agora é como a sequência de aminoácidos de uma proteína é determinada por uma trinca no mRNA? Assim, Crick supôs que existiria um adaptador, ou seja, uma molécula que leva os aminoácidos ao RNA mensageiro, supôs também que este adaptador tem que conter nucleotídeos para reconhecer o RNA mensageiro e ainda que os aminoácidos fossem ligados a cada um dos seus adaptadores por uma enzima. A hipótese do adaptador e da enzima foi confirmada através da descoberta dos RNA transportador (tRNA) e da aminoacil tRNA sintetase respectivamente (WATSON et al., 2015).

Quando se observa o código genético, sugere-se que este evoluiu de modo a minimizar os erros, em razão de mutações na primeira base do código frequentemente resultarem em um aminoácido quimicamente semelhante, além disso, os códigos que possuem pirimidinas na segunda posição especificam aminoácidos hidrofóbicos, as que possuem purinas na segunda posição especificam aminoácidos polares e as transições (A:T por G:C ou G:C por A:T) são as mutações mais frequentes, logo uma mutação na segunda base do códon resultará em um aminoácido semelhante (WATSON et al., *op cit*).

Em 1965 Sidney Brenner, trabalhando com bacteriófagos T4, analisou algumas mutações em um gene que controla a proteína da cabeça do fago, sendo esta proteína mutante formada por uma cadeia polipeptídica mais curta que o tipo selvagem. Comparando as proteínas mutantes e selvagem ele observou que para cada mutante os aminoácidos seguintes que teriam sido inseridos para iniciar a cadeia polipeptídica eram: glutamina, lisina, ácido glutâmico, tirosina, triptofano e serina. Partindo dessa observação Brenner deduziu que alguns códons para esses aminoácidos eram similares e o código destes conseguiam mudar para UAG, assim ele disse que este códon é o de parada (BRENNER; STRETTON; KAPLAN, 1965). Após esta evidencia outros códons de parada foram descobertos como UGA e UAA.

Dessa maneira o código genético tem as seguintes características: é redundante, pois tem alguns aminoácidos que são especificados por mais de uma trinca, não é lido em virgulas, pois, a entrada ou retirada de um nucleotídeo muda todo o quadro de leitura do gene, esta aparentemente organizado de forma que diminua o efeito de uma mutação missence, possui sinais de início e paradas da cadeia polipeptídica, três bases codificam os aminoácidos onde estas tricas são chamadas de códons e cada códon é especificado por apenas um aminoácido.

1.2.2 Moléculas Responsáveis pela Tradução do Código Genético

Para que ocorra a tradução da informação contida no RNA mensageiro precisa-se de uma maquinaria composta de mRNA, tRNA, Aminoacil tRNA sintetase, e ribossomos, responsáveis por traduzir o código genético.

Desde 1957 com os experimentos de Elliot Volkin e Lawrence Astrachan chamado de “pulso ou caça” usando bactérias marcadas por uracila radioativa foi constatado que o RNA se move do núcleo ao citoplasma onde as proteínas são sintetizadas, logo o RNA era considerado um forte candidato a ser o intermediário da transferência de informação do DNA para as proteínas (GRIFFITHS, J.F. Anthony et al. 2015)

O RNA depois de ser transcrito através de uma ação conjunta de várias polimerases, fatores de transcrição, proteínas receptoras, promotores, sequências consenso e sequências fim que dita o término da síntese, este é processado. Depois de ter ocorrido a transcrição do pré-RNA mensageiro tanto em bactérias quanto em eucariontes, constatou-se que em bactérias o pré-RNAm é traduzido ainda quando da etapa de transcrição, este RNA é composto só por regiões codificadoras (éxons), no entanto, em eucariontes o RNAm precisa ser processado para que ocorra a tradução, cujas etapas envolvem: acréscimo do revestimento cap na ponta 5', splicing e acréscimo de nucleotídeos de adenina na extremidade 3' (NELSON .L.David; COX.M.Michael, 2014)

O RNA transportador (tRNA) é composto por 76 nucleotídeos com uma terminação CCA que é universal, é nessa terminação que ocorre a ligação do aminoácido, (JÜHLING et al., 2009). O tRNA tem uma estrutura em forma de trevo com 4 hastes de dupla hélice e três alças unifilamentar, a alça do meio de cada tRNA é chamada de alça anticódon porquê contém uma trinca de nucleotídeos denominada anticódon (ROBERTUS et al., 1974).

O anticódon do tRNA e o códon do RNA mensageiro ligam-se por pareamento de bases e a distância entre a haste aceitadora de aminoácido e a alça do anticódon do tRNA é de 75 Å. Devido a molécula de tRNA possuir essa sequência CCA de cadeia simples bem conservada, foi proposto um modelo sobre a origem do tRNA, o modelo evidencia que o tRNA evoluiu a partir de uma sequência CCA que teve uma função importante na replicação durante o “mundo do RNA”. A hipótese de que os RNA lineares antigos possuem estruturas CCA semelhantes aos tRNA de hoje foi fortalecida devido a gama de processos replicativos que estas sequências CCA do tRNA exerceram na função replicativa (TAMURA, 2015).

Uma das questões mais intrigantes da ciência era saber que tipo de moléculas compunham as células primitivas, Carl Woese (1990) propôs que essas células primitivas tinham uma informação baseada em RNA e um metabolismo não muito eficiente, denominando-o progenotes, anos depois novas pesquisas evidenciaram que o tRNA era responsável por compor essas células e que tinham um anticódon do tipo RNY(R=PURINA,Y=PIRIMIDINA e N= PURINA ou PIRIMIDINA) e a partir destes, foram reconstruídos os tRNA ancestrais, através dessa reconstrução foi proposto que os tRNA foram os primeiros genes. Com a tradução deles, as proteínas produzidas mostraram semelhanças na sua estrutura com várias proteínas de vias metabólicas (DE FARIAS; RÊGO; JOSÉ, 2016b).

A união dessas células primitivas para alguns pesquisadores originou o LUCA (último ancestral comum universal), que foi o primeiro organismo que supostamente continha informação baseada em DNA e um metabolismo eficiente (DE FARIAS; PROSDOCIMI, 2016).

As aminoacil tRNA sintetase são enzimas responsáveis pela a ligação do aminoácido ao seu tRNA. Existem duas classes de aminoacil tRNA sintetase e cada uma leva 10 aminoácidos, entretanto a lisina é o único aminoácido que ocorre tanto na classe 1 quanto na classe 2 (SHEPPARD et al., 2008). Já foi visto que as aminoacil tRNA sintetase de classe 2 são mais antigas que as de classe 1, porque as de classe 2 levam aminoácidos que são produzidos em ambientes prebióticos (KLIPCAN; SAFRO, 2004).

O ribossomo são estruturas responsáveis pela síntese proteica, nele a parte responsável por polimerizar os aminoácidos é composto por RNA e proteínas, o ribossomo possui duas subunidades: uma maior chamada de LSU e uma pequena chamada de SSU. Carl Woese foi o primeiro a estudar as relações evolutivas encontradas nos ribossomos e através deste estudo ele concluiu que existe três domínios da vida: *Archaea*, *Bacteria* e *Eukarya*, essa descoberta foi realizada através de análises filogenéticas das sequências e estruturas secundárias dos RNA ribossômicos da subunidade SSU (WOESE; FOX, 1977; WOESE; KANDLER; WHEELIS, 1990).

O ribossomo configura-se assim como uma maquinaria molecular responsável pela tradução do código genético, com uma interação entre o RNA mensageiro e

RNA transportador, essa estrutura possui um canal longo onde a cadeia do RNA mensageiro passa e três sítios de ligação que interage com o RNA transportador, sítio A (aminoacil), P (peptidil) e E (saída), os sítios A e P são responsáveis pela a execução das ligações peptídicas (NOOR; COYNE, 2006).

A região CCA do RNA transportador liga-se ao aminoácido que está no sítio A, o qual será ligado a outro aminoácido nascente no sítio P do ribossomo através da ligação peptídica, então durante o alongamento as moléculas de RNA transportador localizadas nos três sítios A, P e E atuam em conjunto para translocar o aminoácido do sítio A para o sítio P e depois para o sítio E (NOOR; COYNE, *op cit*).

Embora grande parte das estruturas dos RNA ribossômicos dos três domínios da vida sejam iguais, elas possuem sequências de proteínas que são únicas, uma vez que existem setenta e oito proteínas ribossômicas eucarióticas, sessenta e oito proteínas ribossômicas arqueanas e cinquenta e sete proteínas bacterianas, quando as estruturas primárias destas proteínas são alinhadas vê-se que todas as proteínas arqueanas são homólogas as de eukarya, portanto o ribossomo eucariótico é um ribossomo arqueano (LECOMPTE et al., 2002).

Quando estas sessenta e oito proteínas arqueanas são alinhadas as cinquenta e sete proteínas bacterianas é verificado que só 34 são homólogas entre archaea e bactéria, sendo assim estas 34 proteínas são homologas aos três domínios, entre as 34 proteínas homólogas 15 compõe a subunidade menor e 19 a subunidade maior (SMITH et al., 2008).

Uma série de estudos concluiu que a subunidade maior do ribossomo a LSU é mais antiga que a subunidade menor SSU. Sabendo-se que a formação de peptídeos curtos veio antes da decodificação do RNA mensageiro na subunidade menor, isso quer dizer que os aminoácidos que compõe as proteínas nas subunidades maiores são considerados primordiais. Acredita-se que o ribossomo evoluiu dessa subunidade maior (BOKOV; STEINBERG, 2009; HSIAO et al., 2009; HURY et al., 2006).

O centro da peptidil transferase localizado na subunidade maior do ribossomo é responsável pelas ligações peptídicas, o ribossomo que se conhece hoje evoluiu a partir deste centro em uma série de etapas, a principal foi a formação do sítio P que provavelmente surgiu durante “o mundo do RNA”, com função catalítica,

posteriormente ocorreram novas inserções de moléculas de RNA melhorando ainda mais esse sítio e consequentemente formando o sítio A. A junção do sítio A e P formou o poro de saída do centro da peptidil transferase, depois da formação do sítio P e A as etapas seguintes foram responsáveis por uma melhor estabilidade e reforço do centro peptidil transferase (PETROV et al., 2014).

As proteínas L2, L3, L4, L22 da subunidade maior presentes em todos os domínios da vida, estão próximas ao centro da peptidil transferase, (HSIAO et al., 2013). Experimentos realizados com as proteínas L22 e L4 revelaram que elas reagiram com os centros da peptidil transferase, em particular no local de túnel de saída das proteínas sintetizadas, sendo assim possivelmente elas são responsáveis por compor o túnel de saída das proteínas. A proteína L3 é responsável por fazer o “pisso” da peptidil transferase por que ela reage com RNA ribossômico, que faz o piso da peptidil transferase, já L2 fica localiza do lado oposto da L3 e próxima a L4 (VISHWANATH et al., 2004).

O centro da peptidil transferase é composto unicamente por RNA ribossômico e desprovido de proteínas, foi comprovado que este sítio é muito conservado em todos os domínios da vida, portanto, pode-se considerar que o ribossomo é uma enzima de RNA, porem nas outras partes do ribossomo encontra-se proteínas e estas são responsáveis por deixar o ribossomo mais eficiente em sua catálise (KATUNIN et al., 2002).

O centro da peptidil transferase é composto por 120 nucleotídeos, esta região é responsável por acomodar a estrutura CCA do RNA transportador que leva o aminoácido, neste centro os nucleotídeos se dobram de maneira a constituir um bolso que proporciona um ambiente favorável para formar ligações peptídicas, acredita-se que este bolso se formou através da junção de dois RNAs que se dimerizaram, formando um protoribossoma. Posteriormente, os aminoácidos levados pelo o RNA transportador se acomodaram dentro desse bolso, aumentando a probabilidade de ocorrer ligações peptídicas, estes pequenos peptídeos agiram para melhorar o aparelho sintético ou seja melhorando o protoribossoma (KRUPKIN et al., 2011).

1.3 Evolução do Código Genético

Muitas são as hipóteses sobre a origem do código genético, a primeira veio com Francis Crick chamada de “frozen acidente hypothesis”, que propôs: o código teria sido congelado e evoluiu a partir de um ancestral comum (CRICK, 1968; TAMURA, 2016).

Carl Woese em 1968, propôs a teoria estereoquímica para a origem do código genético, que afirma: a correspondência entre códon e aminoácidos pode acontecer de acordo com a hidrofobicidade dos aminoácidos, pois aminoácidos mais hidrofóbicos conseguem se ligar com mais estabilidade aos códons (WOESE, 1965; WOESE et al., 1966).

Pequenos oligonucleotídeos podem ser formados em ambientes prebióticos sem ajuda de enzimas, e estes conseguem se ligar a aminoácidos através de ligações covalentes, e também podem ocorrer ligações peptídicas quando as interações de vários oligonucleotídeos com seus respectivos aminoácidos ou estas ligações peptídicas podem ser realizadas com ajuda de ribozimas, portanto formando um sistema de codificação primitivo (SELIGMANN; AMZALLAG, 2002).

Um dos principais argumentos contrários a esta teoria diz respeito à estrutura dos tRNA que se conhece hoje, porque este em uma extremidade liga-se a aminoácidos e em outra aos códons, e para alguns pesquisadores essa distância não fortalece o modelo estereoquímico, entretanto já foi proposto que a duplicação dos oligonucleotídeos foi responsável pela formação do tRNA, sendo assim esta evidência fortalece o modelo estereoquímico (SELIGMANN; AMZALLAG, *op cit.*).

O modelo de minimização de erros preconiza que as pressões seletivas fizeram com que o código genético ficasse mais forte com o intuito de minimizar erros nas proteínas que eram sintetizadas, neste modelo o código genético parece está organizado de tal maneira que, quando os nucleotídeos são substituídos os aminoácidos que são mudados possuem características químicas semelhantes aos anteriores, o código genético neste modelo atribui aminoácidos semelhantes com códons semelhantes (ZHU; ZENG; HUANG, 2003). Esta hipótese foi fortalecida através de dados estatísticos por Alff-Steinberger (1969).

Wong em 1975 propôs seu modelo sobre a origem do código genético, neste modelo ele afirma que o código genético evoluiu junto com um sistema bioquímico primitivo, ou seja, é baseado em uma coevolução. O que faz esse modelo ser aceito é uma relação produto – precursor, em que certos aminoácidos que dão origem a outros quimicamente diferentes possuem códons semelhantes (WONG et al., 2016) (Figura 2).

Observa-se na Figura 2 que os aminoácidos com cores iguais possuem a mesma via metabólica em comum, e consequentemente códons semelhantes como é o caso dos aminoácidos prolina, histidina, arginina e ácido glutâmico que possuem a mesma cor, e também serina, arginina, cisteína e triptofano com a mesma cor, e entre outros. A ilustração mostra a relação produto-precursor entre os aminoácidos, pois mesmo possuindo características químicas diferentes eles podem ter origem da mesma via metabólica e códons análogos (WONG, 2005).

Figura 2: Relação produto-precursor dos aminoácidos segundo Wong

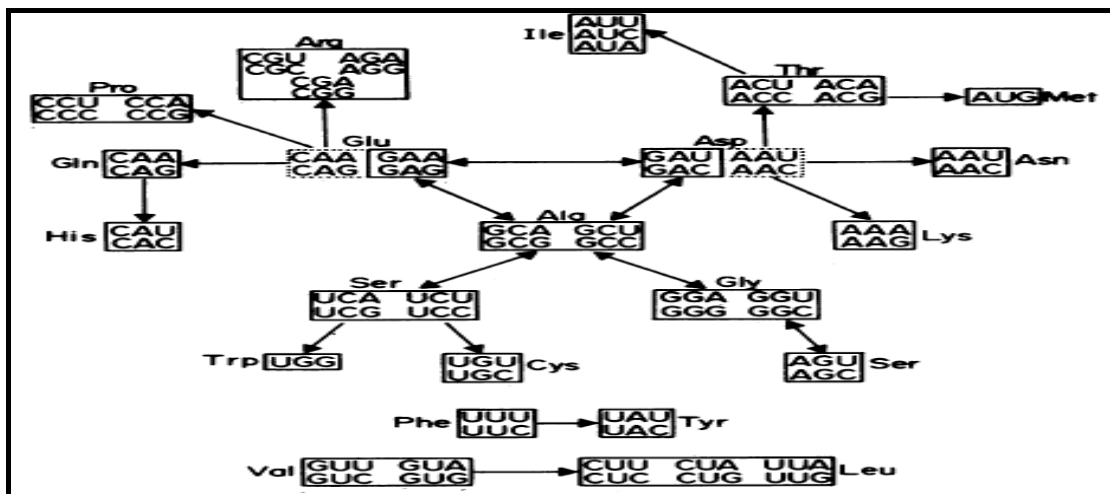
UUU Phe	UCU Ser	UAU Tyr	UGU Cys
UUC Phe	UCC Ser	UAC Tyr	UGC Cys
UUA Leu	UCA Ser	UAA ter	UGA ter/Sec
UUG Leu	UCG Ser	UAG ter/Pyl	UGG Trp
CUU Leu	CCU Pro	CAU His	CCU Arg
CUC Leu	CCC Pro	CAC His	CGC Arg
CUA Leu	CCA Pro	CAA Gln	CGA Arg
CUG Leu	CCG Pro	CAG Gln	CGG Arg
AUU Ile	ACU Thr	AAU Asn	AGU Ser
AUC Ile	ACC Thr	AAC Asn	AGC Ser
AUA Ile	ACA Thr	AAA Lys	AGA Arg
AUG Met	ACG Thr	AAG Lys	AGG Arg
GUU Val	GCU Ala	GAU Asp	GGU Gly
GUC Val	GCC Ala	GAC Asp	GGC Gly
GUA Val	GCA Ala	GAA Glu	GGA Gly
GUG Val	GCG Ala	GAG Glu	GGG Gly

Fonte: Wong, 2005

Wong (2005), afirma ainda que a conversão química do precursor para o produto ocorre enquanto o precursor ainda está ligado a seus RNA adaptadores, então aminoácidos diferentes mais com a mesma via metabólica possuem códons

semelhantes, no seu modelo ele admite que os aminoácidos ácido glutâmico, ácido aspártico, glicina e alanina foram os aminoácidos mais abundantes na atmosfera primitiva, e que estes quatro aminoácidos deram origem a outros 11 aminoácidos. E que os aminoácidos ácido glutâmico e ácido aspártico foram os primeiros a compor o código genético primitivo, pois estes dois aminoácidos possuem códons semelhantes a todos os outros códons dos restantes aminoácidos, e que a partir deles podem ser produzidos diversos aminoácidos como mostra a Figura 3 (WONG, 1975)

Figura 3: Ácido aspártico e ácido glutâmico: Primeiros aminoácidos a compor o código genético primitivo segundo Wong.



Fonte: Wong, 1975.

O modelo de código genético do Hartman é baseado em quatro análises: estruturas da subunidade grande do ribossomo, dos aminoácidos ao redor do centro da peptidil transferase, metabolismo dos aminoácidos primordiais e ligação dos aminoácidos a um código de citosina CG. Em seu modelo este pesquisador afirma que os primeiros aminoácidos a compor o código genético primitivo se ligavam a nucleotídeos CG em um dos braços do RNA transportador primitivo, pois já foi relatado que os aminoácidos tinham capacidade de se ligar a uma haste CG nos RNA transportador, independente do reconhecimento da região do anticódon, e os aminoácidos que se ligavam melhor ao código CG eram aqueles que hoje são

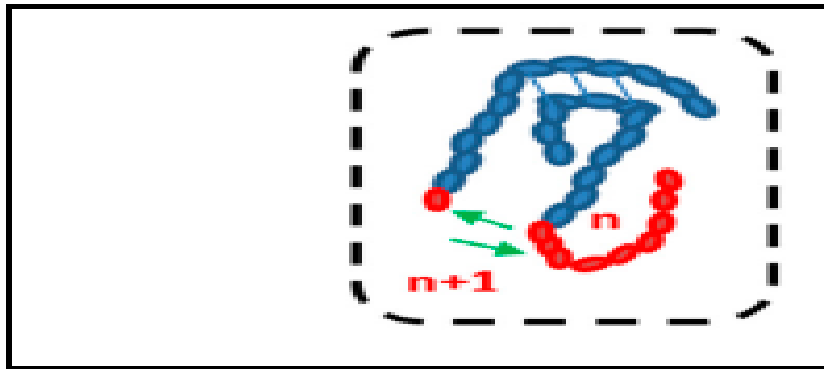
levados por aminoacil sintetase de classe 2, sintetizados mais rapidamente em ambientes prebióticos do que os de classe 1 (HARTMAN, 1995).

Os primeiros aminoácidos codificados eram selecionados por questões metabólicas e pela natureza mais antiga das aminoacil sintetase de classe 2, a partir dessa lógica foi sugerido que os primeiros aminoácidos a se ligarem ao código CG na haste receptora primitiva foram glicina, prolina e alanina, que compartilham códigos CG, posteriormente analisando a subunidade maior do ribossomo que é responsável pela a formação das ligações peptídicas principalmente, nas regiões que ficam próximas ao centro da peptidil transferase foi observado uma abundância de aminoácidos como glicina, lisina e arginina, e em outras análises das proteínas L2, L3, L4 e L22 desta subunidade maior observou-se uma abundância de aminoácidos como glicina, prolina e arginina, portanto foi considerado que os aminoácidos glicina, prolina, alanina, arginina e lisina estão conservados próximo a região do centro da peptidil transferase (PTC), compondo estes aminoácidos sessenta e sete por cento desta região (HARTMAN; SMITH, 2014).

Baseando nestas análises, Hartman considera os aminoácidos glicina, alanina, prolina, arginina e lisina como os primeiros a comporem o código genético primitivo, futuramente com a expansão da adenina vieram lisina, ácido glutâmico, ácido aspártico, glutamina, asparagina, histidina, treonina e serina. Com a introdução da uracila os aminoácidos alanina, isoleucina metionina, fenilalanina, tirosina, triptofano e cisteína, no entanto a questão da arginina e lisina comporem o código genético primitivo é contestada, pelo fato destes aminoácidos requererem muitos passos catalíticos e ser levadas por aminoacil tRNA sintetase de classe 1, deste modo, Hartman defendeu que no começo não havia arginina e lisina mais sim a ornitina, considerada um precursor tanto da arginina quanto da lisina e que a ornitina poderia ser sintetizada a partir de serina em ambientes prebióticos de forma rápida (HARTMAN; SMITH, *op cit.*).

O modelo autoreferente propõe que o código genético surgiu através de ligações entre tRNA, pois estes tRNA ligados com seus respectivos aminoácidos quando juntos facilitava a formação de ligação peptídica, ou seja estes pequenos tRNA tinham a capacidade de dimerizar, atuando como uma peptidil transferase, facilitando a formação de ligações peptídicas como mostrado na Figura 4, onde em vermelho aminoácidos e em azul os tRNA (CARDOSO GUIMARÃES, 2017).

Figura 4: Modelo autoreferente



Fonte: Guimarães, 2017.

Sucessivamente, com a produção de vários peptídeos ou proteínas, estes se ligavam aos próprios dímeros de tRNA melhorando ainda os sistemas, ou seja, estes peptídeos ou proteínas produzidos realizavam feedback positivo estabilizando ainda mais o sistema, os primeiros aminoácidos que se ligaram a estes tRNA segundo o modelo autoreferente foram: glicina, serina e prolina pois estes três aminoácidos possuem características interessantes, como estar presentes em regiões importantes das proteínas, extremidades n terminais, por estar conservados em sítios das aminoacil tRNA sintetase, e no caso da prolina por ser produzida através de ácido glutâmico, abundante em condições prebióticas (GUIMARÃES; MOREIRA; DE FARIAS, 2008).

Em relação a serina e glicina por ter um ciclo serina-glicina que era responsável por produzir estes aminoácidos em ambientes prebióticos, depois com a formação das proteínas estas tinham a capacidade de se ligar a outros tRNA durante sua replicação, tornando mais eficiente a replicação deles, resultando no surgimento dos RNA ribossômicos e RNA mensageiros (GUIMARÃES; MOREIRA; DE FARIAS, *op cit.*).

A formação de um código genético primitivo ainda é um mistério na biologia evolutiva, acredita-se que os tRNA foram as primeiras moléculas a comporem o código genético primitivo, ou seja, foram eles que primeiro integraram um protoribossoma, as junções de várias moléculas de proto-tRNA foram responsáveis por formar o centro da peptidil transferase onde se realizam as ligações peptídicas (DE FARIAS; RÊGO; JOSÉ, 2016a)

2 OBJETIVOS

2.1 Objetivo Geral

Analisar através de modelo estatístico a origem dos primeiros aminoácidos a constituírem o código genético primitivo, destacando aspectos da ancestralidade de proteínas.

2.2 Objetivos Específicos

Avaliar as médias dos aminoácidos derivados da tradução de tRNA ancestrais e dos domínios da vida: *Archaea*, *Bacteria* e *Eukarya* considerando a concentração crescente de aminoácidos nas proteínas;

Avaliar as médias dos aminoácidos derivados das proteínas derivadas da tradução de tRNA ancestrais e domínios da vida *Archaea*, *Bacteria* e *Eukarya* baseado nas classes químicas dos aminoácidos;

Analisar médias e variâncias dos aminoácidos advindos de proteínas que foram traduzidas de tRNA ancestrais e também de todas as proteínas de várias amostras de organismos dos três domínios da vida: *Archaea*, *Bacteria* e *Eukarya*.

3 METODOLOGIA

3.1 Contagem dos Aminoácidos das Proteínas Traduzidas de tRNA Ancestrais

As sequências de tRNA ancestrais foram reconstruídas a partir de análises dos primeiros códigos RNY e depois de obtido os tRNA ancestrais foram traduzidos, apresentando semelhanças com várias proteínas, portanto estas proteínas são consideradas proteínas ancestrais como apresentado na figura 5 (DE FARIAS; RÊGO; JOSÉ, 2016a). A partir de análises destas proteínas foram realizadas contagens dos seus aminoácidos, obtendo-se a frequência de cada aminoácido, usando o Excel e posteriormente calculado a média aritmética e variâncias de cada um dos aminoácidos.

Figura 5: Proteínas advindas da tradução de tRNA ancestrais

Protein	Process	Protein	Process
Elongation factor 1-alfa RNA transport	Translation	Putative ribose/galactose/methyl galactose import	Glycolysis/glycogenesis
Elongation factor 4-mRNA translation assisting	Translation	Glycerate kinase	Glycolysis/glycogenesis
tRNA uridine 5-carboxymethylaminomethyl modification	Translation	Triosephosphate isomerase	Glycolysis/glycogenesis
60S ribosomal protein L3	Translation	Beta-glucosidase A	Glycolysis/glycogenesis
60S ribosomal protein L7a	Translation	Glucose-6-phosphate-1-dehydrogenase	Glycolysis/glycogenesis
60S ribosomal protein L27a-1	Translation	Glucose 6 phosphate isomerase	Glycolysis/glycogenesis
60S ribosomal protein L27a-3	Translation	Phosphoglycerate kinase	Glycolysis/glycogenesis
60S ribosomal protein L27a-2/4	Translation	Glycerol-3-phosphate dehydrogenase	Glycolysis/glycogenesis
50S ribosomal protein L10	Translation	Transketolase	Glycolysis/glycogenesis
Methionyl-tRNA formyl transferase	Translation	α -galactosidase	Glycolysis/glycogenesis
Lysyl-tRNA synthetase	Translation	Diaminopimelate epimerase	Amino acids pathways
Asparaginyl-tRNA synthetase	Translation	L-asparaginase	Amino acids pathways
Glutamyl-tRNA synthetase	Translation	ATP phosphoribosyl transferase	Amino acids pathways
Leucyl-tRNA synthetase	Translation	Histidinol phosphate aminotransferase	Amino acids pathways
Valyl-tRNA synthetase	Translation	4-aminobutyrate aminotransferase	Amino acids pathways
Phenylalanine-tRNA synthetase	Translation	Ornithine carboxylase antizyme	Amino acids pathways
RNA methyltransferase—tRNA modification	Translation	N-acetyl- γ -glutamyl-phosphate reductase	Amino acids pathways
DNA-direct RNA polymerase or RNA-direct RNA polymerase	Transcription	Homoserine kinase	Amino acids pathways
Thymidylate kinase	Nucleotides pathways	Aromatic amino acid aminotransferase	Amino acids pathways
Cytidine deaminase	Nucleotides pathways	Ornithine carbonyltransferase	Amino acids pathways
Uridylate kinase	Nucleotides pathways	Tryptophan synthase α -chain	Amino acids pathways
Orotidine 5-phosphate decarboxilase	Nucleotides pathways	Fatty acid synthase	Lipids pathways
Dihydroorotate dehydrogenase	Nucleotides pathways	CoA mutase	Lipids pathways
Phosphoribosyl formyl glycinamide cyclo-ligase	Nucleotides pathways	Phosphate acyltransferase	Lipids pathways
Phosphoribosyl glycinamide synthase	Nucleotides pathways	Lycopene cyclase	Lipids pathways
		3- β -hydroxysteroid dehydrogenase	Lipids pathways

Fonte: De Farias; Rêgo; José, 2016

3.2 Genomas dos Organismos Domínios da Vida (*Bacteria*, *Archaea* e *Eukarya*).

Foram escolhidos 20 organismos do domínio *Bacteria*, 20 do domínio *Archaea* e 20 do domínio *Eukarya*, sendo adquiridos todos os genomas de cada um deles, obtendo-se a frequência, média aritmética e a variância dos aminoácidos. Os genomas dos organismos foram obtidos do programa online **CODON USAGE DATABASE** (<http://www.kazusa.or.jp/codon/>).

Os organismo referente ao domínio *Bacteria* foram : *Acinetobacter baumannii* 3807, *Bacillus cereus* ATCC10987 5869 CDS's, *Brucella abortus* biovar 1 str. 9-941 3074, *Burkholderia mallei* ATCC 23344 5025, *Clostridium beijerinckii* NCIMB 8052 5020, *Corynebacterium glutamicum* R 3081, *Enterococcus faecalis* V583 3265, *Escherichia coli* 8087, *Klebsiella pneumoniae* subsp *pneumoniae* MGH 78578 5185 CDS's, *Leptospira interrogans* serovar *Lai* str 56601 4725, *Mycobacterium tuberculosis* 4189, *Neisseria meningitidis* 2063, *Prochlorococcus marinus* str. MIT 9313 2265, *Pseudomonas aeruginosa* PA7 6286, *Pseudomonas syringae* pv. *Syringae* B728a 5089, *Rhodococcus* sp *RHA1* 9201, *Staphylococcus aureus* subsp. *Aureus* MW2 4964, *Streptococcus pyogenes* MGAS10270 1987, *Streptomyces coelicolo* A3(2) 8375.

Os organismos referentes ao domínio *Archaea* foram: *Archeaoglobus fugidus* 2407, *Candidatus Methanoregula boonei* 6A8 2450, *Haloarcula marismortui* ATCC 4240, *Halobacterium* sp. *NRC-1* 2605, *Haloquadratum walsbyi* 2646, *Methanobrevibacter smithii* ATCC 35061, *Methanococcus maripaludis* C5 1822, *Methanosaeta thermophila* 1696 CDS's, *Methanosarcina acetivorans* C2A 4540 CDS's, *Methanosarcina mazei* Go1 3378, *Natronomonas pharaonis* DSM 2822, *Pyrobaculum arsenaticum* 13514, *Pyrococcus furiosus* 2108, *Pyrococcus horikoshii* OT3 2067, *Sulfolobus acidocaldarius* DSM 639 2223, *Sulfolobus solfataricus* P2 2994, *Thermococcus kodakarensis* KOD 23 *Thermofilum pendens* Hrk 5 1876, *Thermoplasma volcanium* GSS1 1526 e *Methanoculleus marisnigri* JR1 2489.

Os organismos referentes ao domínio *Eukarya* foram: *Aspergillus oryzae* 12388, *Caenorhabditis elegans* 24994 CDS's, *Candida albicans* 1148, *Cryptococcus neoformans* var. *neoformans* JEC21 6587 CDS's, *Danio rerio* 19062, *Drosophila*

melanogaster 42417, *Gallus gallus* 6017 CDS's, *Glycine max* 1207, *Homo sapiens* 93487 CDS's, *Leishmania major strain Friedlin* 1172 CDS's, *Mus musculus* 53036, *Neurospora crassa* 3953 CDS's, *Oryctolagus cuniculus* 1115 CDS's, *Oryza sativa* 92188, *Rattus norvegicus* 15151 CDS's, *Saccharomyces cerevisiae* 14411, *Schizosaccharomyces pombe* 6109, *Takifugu rubripes* 1119 CDS's e *Xenopus laevis* 13021 CDS's.

3.3 Médias dos tRNA Ancestrais e Domínios da Vida

Depois de obtida as médias dos aminoácidos das proteínas derivadas de tRNA ancestrais e dos domínios da vida, realizou-se a sobreposição das médias, como por exemplo: (aminoácido trp= médias dos tRNA + médias dos domínios /2) em seguida estes resultados foram agrupados na ordem do menor para o maior e construído uma tabela. Posteriormente, para cada aminoácido das proteínas derivadas de tRNA ancestrais e domínios da vida foram produzidas tabelas, relacionando cada um destes aminoácidos com a primeira tabela, selecionando tudo no Excel produzindo em seguida um gráfico de linhas.

3.4: Médias Gerais por Classes Químicas de Aminoácidos

Na construção deste gráfico os aminoácidos foram separados por classes químicas como por exemplo: aminoácidos R apolares alifáticos (tRNA e domínios da vida), grupo R aromáticos (tRNA e domínios da vida), grupo R polares não carregados (tRNA e domínios da vida), grupo R carregado positivamente (tRNA e domínios da vida), grupo R carregado negativamente (tRNA ancestrais e domínios da vida).

Em cada grupo por classe química foi obtido a soma geral das médias tanto nos tRNA ancestrais quanto nos domínios da vida, e posteriormente foram

realizadas as médias entre os tRNA e domínios da vida e em seguida foram divididas por 2, produzindo a tabela no Excel. Com as médias dos tRNA e domínios da vida foram produzidas uma tabela para cada um, depois relacionado as classes químicas dos aminoácidos com a primeira tabela e por último selecionado tudo e realizado um gráfico de barra no Excel.

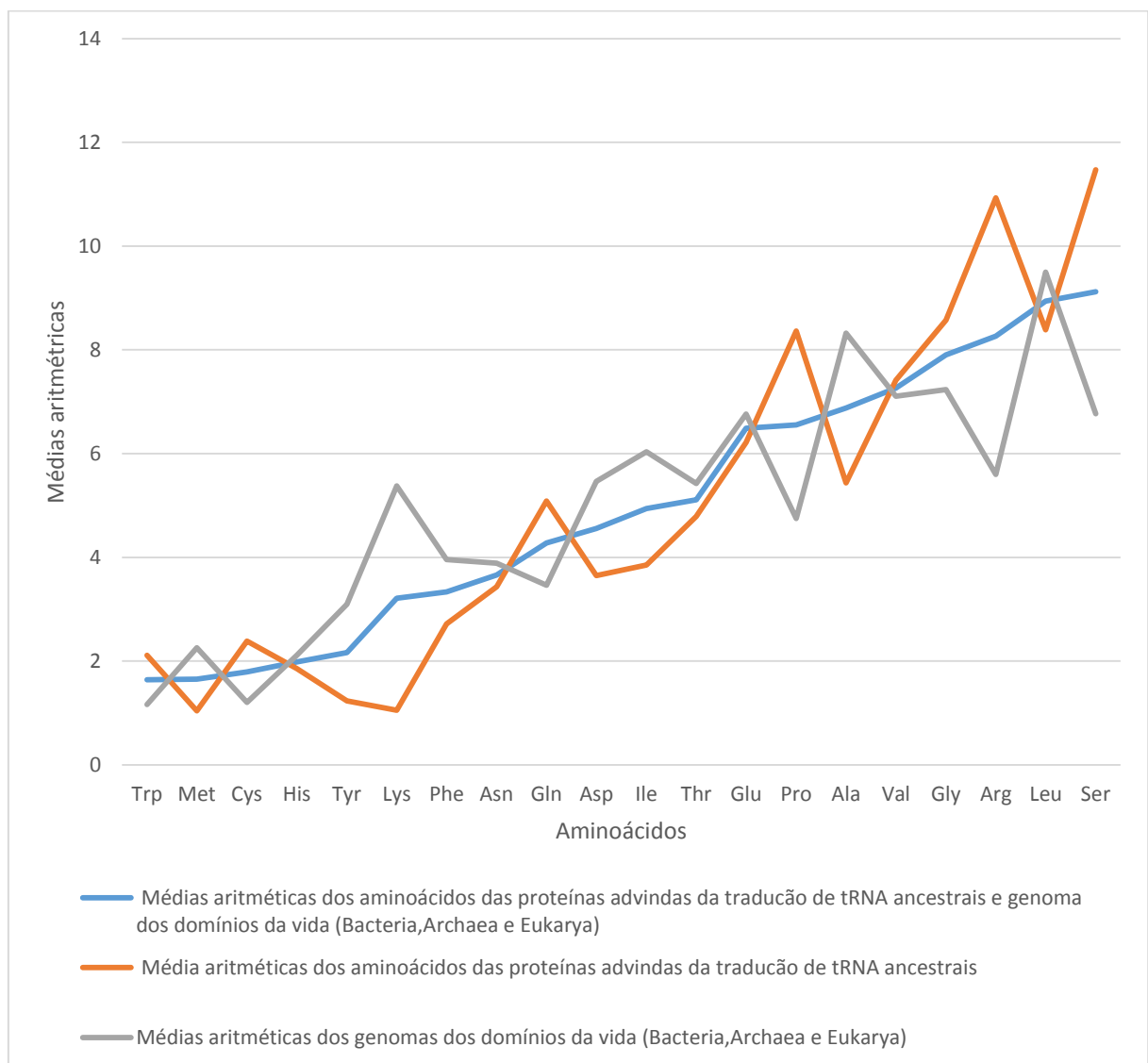
3.5 Análises das Variâncias

Para a construção do gráfico das variâncias foi usado a variâncias dos aminoácidos das proteínas derivadas da tradução de tRNA ancestrais e dos domínios da vida, disponibilizando os dados em uma tabela, posteriormente foi colocado uma tabela com a médias dos aminoácidos das proteínas derivadas da tradução de tRNA ancestrais e outra com os domínios da vida e relacionando essas duas com a primeira tabela, produzindo um gráfico de linhas no Excel.

4 RESULTADOS

No Gráfico 1 são apresentadas as médias aritméticas dos aminoácidos advindos da tradução dos tRNA ancestrais e dos domínios da vida, considerando os níveis crescentes de concentração dos aminoácidos nas proteínas. E ainda as médias aritméticas da sobreposição destes dois parâmetros.

Gráfico 1: Médias dos aminoácidos advindos de proteínas traduzidas de tRNA ancestrais e dos três domínios da vida



Fonte: Elaborado pelo autor, 2018

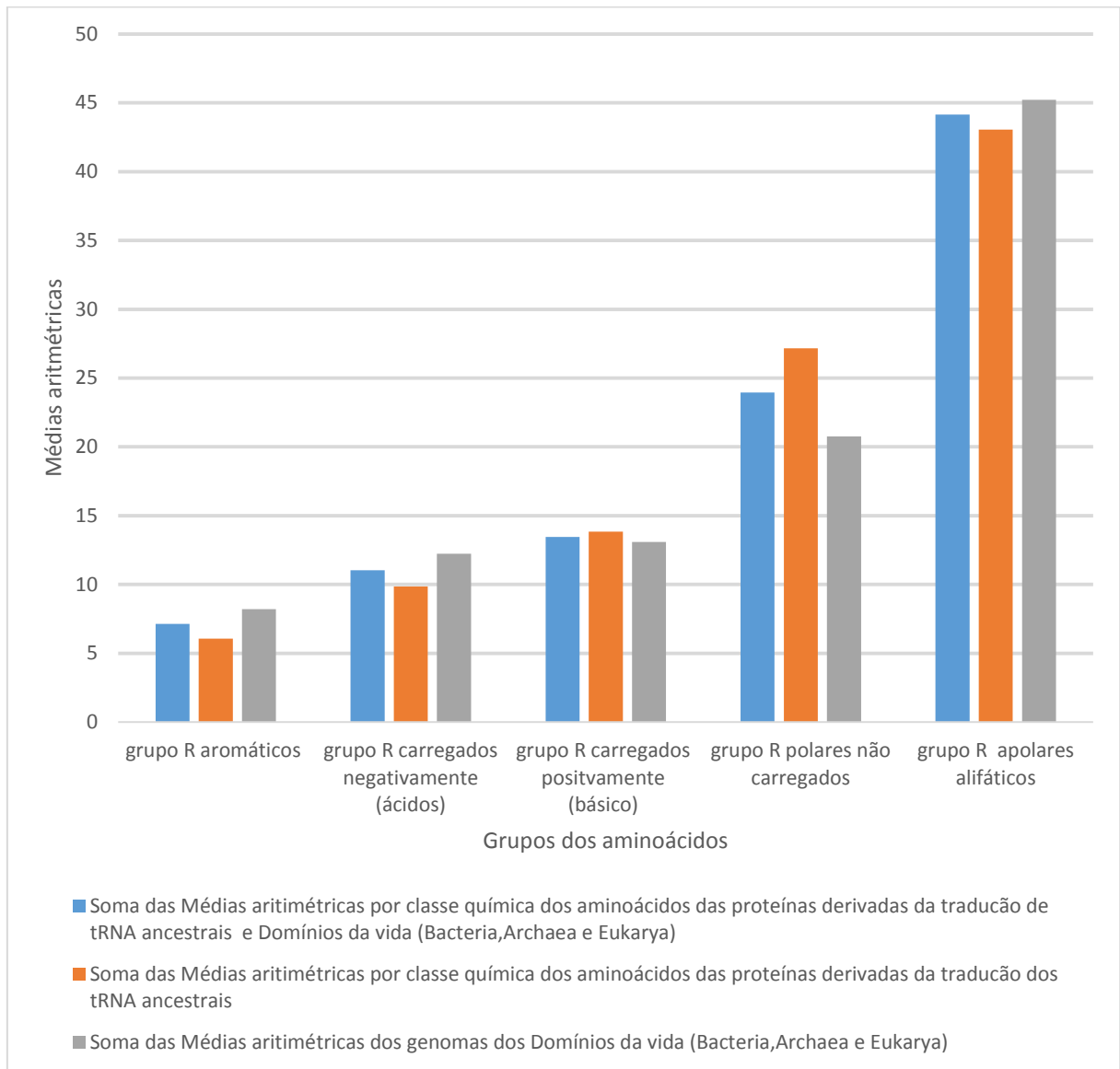
Os dados evidenciados no Gráfico 1 revelam um perfil diferente para o aminoácido não prebiótico lisina, em relação aos demais aminoácidos que o precedem na escala ascendente de concentração de aminoácidos, considerando os parâmetros analisados. A lisina apresenta um pico de concentração do genoma dos domínios da vida e uma baixa concentração relacionada as proteínas advindas da tradução de tRNA ancestrais. Tendo em vista as médias da sobreposição dos parâmetros (domínios da vida e proteínas advindas da tradução dos tRNA), o padrão comportamental ascendente da sobreposição se assemelha ao padrão ascendente das médias dos domínios da vida.

Aminoácidos prebióticos como isoleucina, ácido aspártico e alanina, embora registrem médias de concentração maiores que a lisina, revelam comportamento semelhante a este, ou seja, baixa concentração nos tRNA ancestrais e alta concentração nos domínios da vida, contudo, o padrão comportamental da sobreposição para os aminoácidos prebióticos supracitados revelam um perfil linear com relação aos domínios da vida (Gráfico 1).

Os aminoácidos arginina e glutamina, não prebióticos, assim como, os aminoácidos prolina e serina, prebióticos, apresentam padrão comportamental semelhante com relação aos aminoácidos advindos das proteínas traduzidas de tRNA ancestrais, revelando um aumento de concentração neste parâmetro. No que diz respeito aos aminoácidos do domínio da vida, arginina, glutamina, prolina e serina apresentam médias de concentração mais baixas. Destaca-se ainda acentuado aumento de concentração dos aminoácidos arginina e serina com registro de médias altas nas proteínas advindas da tradução de tRNA ancestrais (Gráfico 1).

Entre os aminoácidos que apresentaram as mais baixas concentrações nos parâmetros analisados, destacam-se triptofano, metionina, cisteína e histidina, todos não prebióticos (Gráfico 1).

Gráfico 2: Médias aritméticas dos aminoácidos por classe química: tRNA ancestrais e domínios da vida



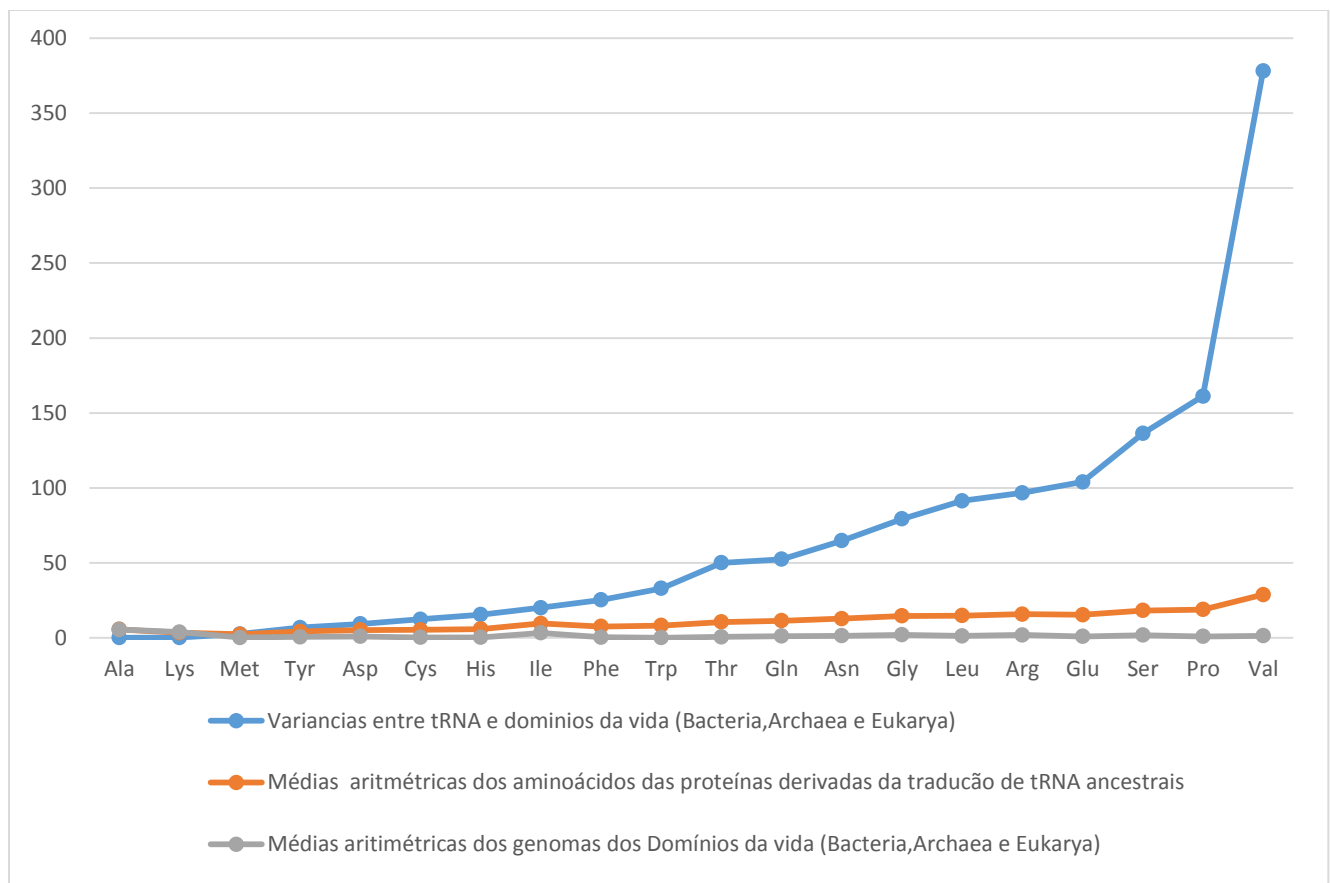
Fonte: Elaborado pelo autor, 2018

Encontra-se evidenciados no Gráfico 2 os resultados dos padrões de distribuição dos aminoácidos de acordo com as classes químicas, considerando as médias dos parâmetros aminoácidos advindos da tradução de tRNA ancestrais, domínios da vida e a sobreposição destes dois parâmetros.

Os dados apresentados no Gráfico 2 revelam uma alta concentração de aminoácidos apolares alifáticos, com valores médios que variaram entre 40 a 45. Constituem o grupo de aminoácidos apolares alifáticos a metionina, isoleucina,

leucina, valina, prolina, alanina e glicina. Não se observa, contudo, um padrão diferencial significativo com relação as médias entre os parâmetros analisados e a sobreposição dos mesmos, tanto para o grupo dos aminoácidos apolares, quanto para os demais grupos.

Gráfico 3: Variâncias: tRNA ancestrais e domínios da vida



Fonte: Elaborado pelo autor, 2018

Os dados apresentados no Gráfico 3 mostram a análise das variâncias dos aminoácidos derivados da tradução de tRNA ancestrais e domínios da vida. Estes resultados revelam que, para os aminoácidos fenilalanina, triptofano, treonina, glutamina, asparagina, glicina, leucina, arginina, ácido glutâmico, serina, prolina e valina os índices de variância foram crescentes e com padrão comportamental apresentando um distanciamento das médias dos parâmetros analisados. Para os demais aminoácidos avaliados não se observa um distanciamento dos índices de variância e as médias dos parâmetros

5 DISCUSSÃO

Os dados apresentados no Gráfico 1 relativos ao aminoácido lisina indicam valores medianos altos dos aminoácidos que compõem os domínios da vida. O aumento da concentração de lisina, aminoácido não prebiótico, pode estar relacionado ao fato deste ter sido produto de uma mutação de guanina (G), para adenina (A) na segunda base do códon do aminoácido arginina, no começo da formação do código genético isso implicou em um mecanismo de erro positivo no código, uma vez que, tais mutações resultaram na produção de lisina, (VAN DER LINDEN; DE FARIAS, 2006).

Estudos anteriores realizados por Hartman; Smith, (2014), comprovam que aminoácidos básicos como argina e lisina, estavam em seleção no começo do sistema de tradução primitivo pois eles iriam contribuir para um maior comprimento da cadeia lateral, consequentemente formando alfa hélices nas proteínas sintetizadas. Estes dados são corroborados pelos resultados apresentados no Gráfico 1, com relação ao aminoácido arginina que apresenta elevadas médias de aminoácidos advindos da tradução de tRNA ancestrais.

Hartman; Smith, (op *cit.*) propuseram que os aminoácidos arginina e lisina no começo da formação do código genético, eram produzidos a partir de um precursor conhecido como ornitina que é sintetizado através de serina por meio de um ciclo do ácido cítrico incompleto, evidência confirmada pelo fato do código de serina AGC ser compartilhado com AGG e AGA de ornitina, portanto pode-se inferir que o aumento de arginina nas proteínas derivadas de tRNA ancestrais pode ser explicado pelo aumento de serina. Consequentemente, o primeiro código de serina seria AGG ou AGA, com a introdução da base uracila em serina, possivelmente esta base passasse a compor a estrutura do códon da arginina.

Uma das contestações em relação a lisina e arginina a comporem o código genético primitivo é que estes não são produzidos em ambientes prebióticos segundo Wong et al., (2016).

O fato das médias de arginina encontrarem-se altas nas proteínas derivadas de tRNA ancestrais e estas médias mostrarem-se baixas em lisina, bem como as

proteínas advindas dos domínios da vida apresentarem-se altas, pode-se inferir que arginina foi o primeiro aminoácido básico a ser incorporado no início da formação do código genético primitivo, e com as mutações na segunda base da arginina durante o processo evolutivo, lisina começou a crescer suas médias nas proteínas dos domínios da vida, contribuindo desta maneira para formação de hélices alfas nas proteínas e para a termoestabilidade do sistema biológico.

Os aminoácidos glicina, alanina, serina, ácido aspártico, ácido glutâmico, valina, leucina, isoleucina, prolina e treonina, todos prebióticos, que são chamados de fase 1, provavelmente apresentaram uma maior capacidade de formação em ambientes inóspitos (WONG, 2005), como aqueles simulados por Miller e Urey na atmosfera primitiva (LAZCANO; BADA, 2003).

Baseado nestas informações pode-se supor que esses aminoácidos foram os primeiros a compor o código genético primitivo. Como mostrado no Gráfico 1, as proteínas derivadas da tradução de tRNA ancestrais nos aminoácidos glicina, serina e prolina apresentaram médias elevadas para este parâmetro. Segundo Cardoso Guimaraes (2017) poderiam ser os primeiros a compor o código genético primitivo, uma vez que, os mesmos possuem certas características como: estar presentes em regiões importantes das proteínas como extremidades n terminais, estar conservados em sítios das aminoacil tRNA sintetase, e no caso da prolina por ser produzido através de ácido glutâmico, abundante em condições prebióticas e por também ter um ciclo serina-glicina responsável por produzir serina e glicina em condições prebiótica.

Hartman; Smith (2014) também defendem a presença de glicina e prolina como componentes do código genético primitivo, estes aminoácidos encontram-se em abundância em proteínas próxima ao centro da peptidil transferase, portanto supõe-se que glicina, prolina e serina apresentam fortes evidências na composição do código primitivo.

Já para Wong (1975), os primeiros aminoácidos a constituírem o código genético foram, ácido glutâmico e ácido aspártico pelo fato destes compartilharem vários códons com outros aminoácidos e ainda por estarem presentes em ambientes prebióticos.

Entretanto Hartman; Smith (2014), admitem que estes aminoácidos não poderiam ser os primeiros a comporem o código genético, já que através de análises de proteínas ribossomais eles concluíram que no início da formação do código genético as primeiras proteínas a se ligarem no sistema teriam a composição de aminoácidos iguais as das proteínas ribossomais. Partindo deste pressuposto, estes pesquisadores constataram que ácido glutâmico e ácido aspártico possuem grupamentos eletricamente carregados, portanto, estas primeiras proteínas que se ligariam ao RNA se elas possuísem aminoácidos negativamente carregados não conseguiriam se ligar ao RNA em razão deste possuir um grupamento fosfato negativamente carregado.

Considerando o aminoácido alanina, com média baixa nas proteínas derivadas de tRNA ancestrais e alta média nas proteínas dos domínios da vida (Gráfico 1), Hartman; Smith (2014) afirmam através de análises de proteínas próximas ao sitio da peptidil transferase, que alanina se encontra bem conservada nestas proteínas e Wong et al., (2016) corroboram com a ideia de que este aminoácido poderia ser produzido em condições prebióticas.

Já Cardoso Guimaraes (2017), considera o aminoácido glicina com maior probabilidade de ser mais facilmente produzido em condições prebióticas, deste modo, pode-se inferir que no início da formação do código genético primitivo glicina estaria compondo o código genético (Gráfico 1), das proteínas derivadas de tRNA ancestrais e naquelas condições ser produzida em quantidades suficientes para sua incorporação ao código genético.

A evidência da presença do aminoácido glicina compondo o código genético primitivo, pelo fato deste aminoácido ser bem hidrofóbico é sustentada pelos estudos de Woese et al., (1966). Carater relevante na formação do sistema de tradução de proteínas primitivas, assim esta pesquisa propõe a que alta concentração de alanina nas proteínas dos dominios da vida (*Arqueia*, *Eukarya* e *Bacteria*) se deve ao fato dela substituir glicina, pois no decorrer da evolução do código genético as mudanças teriam que ocorrer produzindo substituições da glicina por alanina.

Assim sendo, o código ainda continuaria com um aminoácido semelhante na sua composição em relação ao fator da hidrofobicidade, implicando em melhorias em outros adicionais no código genético primitivo, porque as proteínas sintetizadas por

ele, agora apresentariam alanina na sua composição, contribuindo sobremaneira para a formação de alfa hélices nas proteínas, em comparação com a arginina pois sabe-se que a alanina é o aminoácido que mais contribui com alfa hélices dentre todos os demais.

O caráter hidrofobicidade analisado em pesquisas desenvolvidas por Woese, (1965); Woese et al., (1966), preconizam que na formação do sistema de tradução primitivo de proteínas a interação de aminoácidos e nucleotídeos foi de extrema importância e que os aminoácidos mais hidrofóbicos foram os primeiros a consolidar essa relação entre os nucleotídeos.

Os dados apresentados no Gráfico 2 mostram que as médias totais dos aminoácidos tanto derivados das proteínas traduzidas de tRNA ancestrais quanto de domínios da vida, apresentaram maiores valores para os aminoácidos hidrofóbicos, portanto pode-se inferir que a hidrofobicidade se constituiu como um fator de seleção para introdução dos aminoácidos no código genético primitivo, resultados estes que fortalecem o modelo estereoquímico proposto por Woese, (1965).

Ainda relacionado ao caráter de hidrofobicidade das proteínas, destaca-se que os aminoácidos hidrofóbicos foram e continuam sendo importantes na estrutura interna das proteínas, uma vez que, sem os núcleos hidrofóbicos de aminoácidos nas proteínas não poderia ter-se originado a vida, pelo fato destes núcleos hidrofóbicos serem os responsáveis pela estabilização das estruturas proteicas durante o processo evolutivo.

Os dados da variância apresentados no Gráfico 3, mostram que os aminoácidos fenilalanina, triptofano, treonina, glutamina, asparagina, glicina, leucina, arginina, ácido glutâmico, serina, prolina e valina apresentaram maiores índices de variância entre as médias dos parâmetros analisados.

É crível supor que estes aminoácidos foram alvos de seleção para compor o código genético primitivo, por possuírem características químicas que favoreceram aos mesmos serem possíveis alvos de seleção (WONG, 1975; ROMEU, 2017; WOESE, 1965; HARTMAN; SMITH, 2014).

Considerando, os aminoácidos arginina, ácido glutâmico, serina, prolina e valina, com os maiores índices de variância, supõe-se que estes se constituíram nos

primeiros aminoácidos a entrarem no código genético primitivo e consequentemente, passaram a estimular outros aminoácidos a se incorporem ao código, ou seja, estes aminoácidos iniciais funcionaram como um "start" para que outros aminoácidos mais hidrofóbicos e com características de aminoácidos de fase 1 entrassem no código genético.

Em relação aos aminoácidos alanina, asparagina, cisteína, isoleucina, histidina, lisina e tirosina ambos apresentaram baixos índices de variâncias apresentados no gráfico 3 . Sendo assim podemos supor que estes aminoácidos só foram alvo de seleção pelo o código genético posteriormente através das substituições daqueles que entraram primeiro no código genético primitivo argumentado por (WONG, 1975; ROMEU, 2017; WOESE, 1965; HARTMAN; SMITH, 2014).

Assim sendo, os autores desta pesquisa entendem que no início da formação do sistema de tradução primitivo os aminoácidos serina, glicina prolina, alanina e arginina foram os aqueles que primeiro entraram no código genético.

É fato que quando o código genético primitivo surgiu, este tinha um repertório de aminoácidos muito baixo em comparação com o atual e por consequência as proteínas sintetizadas por ele vinham com poucos aminoácidos em sua composição, ou seja, as proteínas primitivas eram pequenas e com número reduzido de aminoácidos.

Experimentos realizados por Shibue et al., (2018) com a proteína NDK (nucleosídeo difosfato kinase), uma proteína que possivelmente estava no ancestral de *archaeae* e *bacteria*, os pesquisadores criaram variantes desta proteína retirando aminoácidos da NDK, submetendo estas variantes a testes de termoestabilidade e atividade catalítica. Estas variantes apresentaram 7 aminoácidos dos 10 que foram encontrados nos experimentos de Miller e Urey, dentre estes estavam alanina, serina, ácido aspártico, ácido glutâmico, valina, prolina e leucina, além dos aminoácidos asparagina, arginina, tirosina e histidina, considerados não prebióticos.

Durante a produção destas variantes foram retirados os aminoácidos cisteína, fenilalanina, glutamina, treonina, triptofano, isoleucina e serina, em seguida foram testadas as suas termoestabilidade e atividade catalítica, obtendo-se como resultado uma variante que permaneceu com boa estabilidade e atividade catalítica,

sendo assim, os aminoácidos que são produzidos em condições prebióticas que compõe essas proteínas contribuem para sua atividade termoestável e catalítica, entretanto os aminoácidos como asparagina, arginina, tirosina, e histidina tidos como não prebióticos, também contribuíram para a atividade desta proteína, portanto, para os autores da pesquisa não se pode descartar a possibilidade destes aminoácidos serem produzidos em ambientes prebióticos (SHIBUE, et al., *op cit.*).

Outros autores, contudo, admitem que estes aminoácidos só vieram adentrar nas proteínas muito posteriormente, pois eles contribuem melhor para a atividade catalítica. Análises mais acuradas revelaram que os aminoácidos prebióticos presentes nas variantes contribuíram mais para a estabilidade da proteína do que para sua atividade catalítica, logo foi verificado que aminoácidos como serina e treonina são bem conservados em certas estruturas, importantes da NDK e sabendo que os aminoácidos que são produzidos em ambientes prebióticos contribuíram para uma melhor estabilidade, as proteínas primordiais possivelmente eram compostas por um conjunto limitado de aminoácidos, ou seja, não estavam diretamente envolvidas na catálise (SHIBUE et al., 2018).

Para os pesquisadores Keller et al., (2017)) e Varma et al., (2018), entretanto, as primeiras atividades catalíticas estiveram associadas a metais e cofatores como, ferro e moléculas contendo enxofre em sua composição, atuando sobre substratos intermediários do ciclo do ácido cítrico.

Comparando os resultados desta pesquisa com os dados de Shibue et al., (2018), pode-se concluir que os aminoácidos serina, leucina, prolina, treonina, arginina, valina, ácido glutâmico, ácido aspártico e alanina que estão compondo a NDK são os mesmos aminoácidos que apresentaram valores de médias altas expostas no Gráfico 1, especialmente nos aminoácidos das proteínas derivadas da tradução de tRNA ancestrais.

Muitos dos aminoácidos da NDK são R apolares alifáticos, destacados no Gráfico 2, onde se observou as maiores médias nos aminoácidos do tipo R apolares alifáticos, ou seja, aqueles que mais contribuíram para a estabilidade proteica.

Por fim, quando comparados os aminoácidos da NDK com os resultados expressos no Gráfico 3, que apresentado os resultados das análises de variância constata-se que, grande parte dos aminoácidos da NDK são encontrados nos dados

do referido Gráfico, portanto, estes aminoácidos provavelmente estavam sendo alvos de seleção pelo o código genético primitivo, e por conseguinte foram estes aminoácidos que também fizeram parte da possível proteína ancestral, que não tinha uma função necessariamente catalítica mais sim de auxiliar em reações químicas, devido esta possuir em grande número aminoácidos que exercia mais função de estabilidade que catalítica.

Em resumo, analisando toda a discussão entende-se que o código genético foi moldado por forte seleção natural. A seleção natural age randomicamente e não sabe o vai produzir, faz uso de estruturas já existentes, ou seja, um sistema quando formado com uma certa função no decorrer da evolução produz novos sistemas mais elaborados partindo de sistemas pré existentes (KAPOOR et al., 2016).

Neste contexto, a pesquisa ora apresentada entende que os aminoácidos a comporem parte da primeira proteína não tinham uma função necessariamente catalítica e sim de estabilidade proteica, aminoácidos como serina, glicina e prolina, que segundo o modelo autoreferente constituíram-se nos primeiros a comporem o código genético primitivo, poderiam ter formado as primeiras proteínas, já que estes aminoácidos contribuem para uma maior estabilidade de proteínas, ou seja, as proteínas que atuavam realizando feedback positivos eram estáveis e pouco catalíticas, deixando o código primitivo mais robusto e eficiente.

Com o decorrer do processo evolutivo os aminoácidos que estavam adentrado no código genético primitivo com mais funções catalíticas como arginina, histidina, asparagina e tirosina conseqüentemente, fizeram parte destas proteínas, e posteriormente começaram a atuar em vias metabólicas com função enzimática, portanto as primeiras proteínas que atuavam em vias metabólicas de modo enzimático eram proteínas que evoluíram de uma linhagem de proteínas que atuavam realizando o feedback positivo.

6 CONCLUSÕES

Tomando como base análises estatísticas de aminoácidos de proteínas derivadas de tRNA ancestrais e dos genomas dos três domínios da vida (*Bacteria*, *Archaea* e *Eukarya*), o aminoácido lisina apresentou um comportamento que oscilou entre os parâmetros domínios da vida (alto) e nas proteínas derivadas de tRNA ancestrais (baixo);

Os aminoácidos arginina e serina mostraram altas médias para os tRNA ancestrais, evidenciando que o aumento na concentração de serina explica o aumento na concentração de arginina. A serina em condições prebióticas atuava como precursor da ornitina, que por sua vez, produzia arginina e lisina;

Entre as análises das classes químicas dos aminoácidos observou-se que as maiores médias estavam na classe dos grupos R apolares alifáticos. Tais análises fortalecem o modelo estereoquímico do código genético, que preconiza que os primeiros aminoácidos a constituírem o código provavelmente, apresentavam maiores índices de hidrofobicidade;

Os altos níveis de alanina nos domínios da vida devem-se as substituições que ocorreram no processo evolutivo do aminoácido glicina por alanina, mantendo-se a hidrofobicidade e contribuindo, sobremaneira, com a formação de alfas hélices nas proteínas;

Os aminoácidos que apresentaram maiores índices de variância, possivelmente foram alvos de seleção para compor o código genético primitivo. Entre estes: arginina, ácido glutâmico, serina, prolina e valina podem ter atuando como “start” para aminoácidos mais hidrofóbicos e da fase 1 entrassem no código genético;

Na evolução do código genético, proteínas com aminoácidos que conferissem maior estabilidade na estrutura de proteínas ancestrais desenvolveram maior adaptabilidade, quando comparadas aquelas com capacidade catalítica, favorecendo a diminuição da entropia nos ambientes em evolução.

REFERÊNCIAS

ALFF-STEINBERGER, C. the Genetic Code and Error Transmission. **Proceedings of the National Academy of Sciences**, v. 64, n. 2, p. 584–591, 1969.

BARKS, H. L. et al. Guanine, adenine, and hypoxanthine production in UV-irradiated formamide solutions: Relaxation of the requirements for prebiotic purine nucleobase formation. **ChemBioChem**, v. 11, n. 9, p. 1240–1243, 2010.

BOKOV, K.; STEINBERG, S. V. A hierarchical model for evolution of 23S ribosomal RNA. **Nature**, v. 457, n. 7232, p. 977–980, 2009.

BRENNER, S.; STRETTON, A. O. W.; KAPLAN, S. Genetic code: The “nonsense” triplets for chain termination and their suppression. **Nature**, v. 206, n. 4988, p. 994–998, 1965.

CARDOSO GUIMARÃES, R. Self-Referential Encoding on Modules of Anticodon Pairs—Roots of the Biological Flow System. **Life**, v. 7, n. 2, p. 16, 2017.

CHYBA, C. et al. Cometary delivery of organic molecules to the early Earth. **Science**, v. 249, n. 4967, p. 366–373, 1990.

CIVIŠ, S. et al. Amino acid formation induced by high-power laser in CO₂/CO- N₂-H₂O gas mixtures. **Chemical Physics Letters**, v. 386, n. 1–3, p. 169–173, 2004.

CRICK, FRANCIS. BRENNER, S. crick192(2).pdf. **Nature**, v. 1227, p. 6, 1961.

CRICK, F. H. C. The origin of the genetic code. **Journal of Molecular Biology**, v. 38, n. 3, p. 367–379, 1968.

DE FARIAS, S.; RÊGO, T.; JOSÉ, M. tRNA Core Hypothesis for the Transition from the RNA World to the Ribonucleoprotein World. **Life**, v. 6, n. 2, p. 15, 2016a.

DE FARIAS, S. T.; PROSDOCIMI, F. Buds of the tree: the highway to the last universal common ancestor. **International Journal of Astrobiology**, p. 1–9, 2016.

DE FARIAS, S. T.; RÊGO, T. G.; JOSÉ, M. V. A proposal of the proteome before the last universal common ancestor (LUCA). **International Journal of Astrobiology**, v.

15, n. 1, p. 27–31, 2016b.

DWORKIN, J. P.; LAZCANO, A.; MILLER, S. L. The roads to and from the RNA world. **Journal of Theoretical Biology**, v. 222, n. 1, p. 127–134, 2003.

FERUS, M. et al. Formation of nucleobases in a Miller–Urey reducing atmosphere. **Proceedings of the National Academy of Sciences**, v. 114, n. 17, p. 4306–4311, 2017.

GRIFFITHS, J. F. Anthony. et al. 2015, **Introdução á Genética**. Tradução: Idilia Vanzelloti... 10 ed Rio de Janeiro: Guanabara Koogan, 2015

GUIMARÃES, R. C.; MOREIRA, C. H. C.; DE FARIAS, S. T. A self-referential model for the formation of the genetic code. **Theory in Biosciences**, v. 127, n. 3, p. 249–270, 2008.

HARTMAN, H. Speculations on the origin of the genetic code. **Journal of Molecular Evolution**, v. 40, n. 5, p. 541–544, 1995.

HARTMAN, H.; SMITH, T. The Evolution of the Ribosome and the Genetic Code. **Life**, v. 4, n. 2, p. 227–249, 2014.

HSIAO, C. et al. Peeling the onion: Ribosomes are ancient molecular fossils. **Molecular Biology and Evolution**, v. 26, n. 11, p. 2415–2425, 2009.

HSIAO, C. et al. Molecular paleontology: A biochemical model of the ancestral ribosome. **Nucleic Acids Research**, v. 41, n. 5, p. 3373–3385, 2013.

HURY, J. et al. Ribosome origins: The relative age of 23S rRNA Domains. **Origins of Life and Evolution of the Biosphere**, v. 36, n. 4, p. 421–429, 2006.

JÜHLING, F. et al. tRNAdb 2009: Compilation of tRNA sequences and tRNA genes. **Nucleic Acids Research**, v. 37, n. SUPPL. 1, p. 159–162, 2009.

KAPOOR, T. M. et al. American Association for the Advancement of Science. v. 311, n. 5759, p. 388–391, 2016.

KATUNIN, V. I. et al. Coupling of GTP Hydrolysis by Elongation Factor G to Translocation and Factor Recycling on the Ribosome Coupling of GTP Hydrolysis by Elongation Factor G to Translocation and Factor Recycling on the Ribosome †. **Society**, p. 12806–12812, 2002.

KAUFFMAN, S. A. Approaches to the Origin of Life on Earth. **Life**, v. 1, n. 1, p. 34–48, 2011.

KELLER, M. A. et al. Sulfate radicals enable a non-enzymatic Krebs cycle precursor. **Nature Ecology and Evolution**, v. 1, n. 4, p. 1–9, 2017.

KLIPCAN, L.; SAFRO, M. Amino acid biogenesis, evolution of the genetic code and aminoacyl-tRNA synthetases. **Journal of Theoretical Biology**, v. 228, n. 3, p. 389–396, 2004.

KRUPKIN, M. et al. A vestige of a prebiotic bonding machine is functioning within the contemporary ribosome. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 366, n. 1580, p. 2972–2978, 2011.

LAZCANO, A.; BADA, J. L. THE 1953 STANLEY L. MILLER EXPERIMENT : p. 235–242, 2003.

LECOMPTE, O. et al. Comparative analysis of ribosomal proteins in complete genomes: An example of reductive evolution at the domain scale. **Nucleic Acids Research**, v. 30, n. 24, p. 5382–5390, 2002.

MARTIN, L.; UNRAU, P.; MÜLLER, U. RNA Synthesis by in Vitro Selected Ribozymes for Recreating an RNA World. **Life**, v. 5, n. 1, p. 247–268, 2015.

MCKENNEY, K.; ALFONZO, J. From Prebiotics to Probiotics: The Evolution and Functions of tRNA Modifications. **Life**, v. 6, n. 1, p. 13, 2016.

NELSON.L.David;COX.M.Michael,**Princípios de Bioquímica de Lehninger**.Tradução :Ana Beatris Gorini da Viega... et al .6 ed. Porto Alegre:artmed,2014.

NIRENBERG, M. W.; MATTHAEI, J. H. H. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. **Proceedings of the National Academy of Sciences of the United States of America**, v. 47, n. 10, p. 1588–602, 1961.

NOOR, M. A F.; COYNE, J. A. Speciation in the New Millennium : What $\hat{\epsilon}^{\text{TM}}$ s Left to Know ? **Israel journal of ecology and evolution**, v. 52, n. April 2014, p. 431–441, 2006.

PETROV, A. S. et al. Evolution of the ribosome at atomic resolution. **Proceedings of the National Academy of Sciences**, v. 111, n. 28, p. 10251–10256, 2014.

PINO, S. et al. From Formamide to RNA, the Path Is Tenuous but Continuous. **Life**, v. 5, n. 1, p. 372–384, 2015.

ROBERTUS, J. D. et al. Structure of yeast phenylalanine tRNA at 3 Å resolution. **Nature**, v. 250, n. 5467, p. 546–551, 1974.

SELIGMANN, H.; AMZALLAG, G. N. Chemical interactions between amino acid and RNA: multiplicity of the levels of specificity explains origin of the genetic code. **Die Naturwissenschaften**, v. 89, n. 12, p. 542–551, 2002.

SHABALINA, S. A.; OGURTSOV, A. Y.; SPIRIDONOV, N. A. A periodic pattern of mRNA secondary structure created by the genetic code. **Nucleic Acids Research**, v. 34, n. 8, p. 2428–2437, 2006.

SHEPPARD, K. et al. From one amino acid to another: tRNA-dependent amino acid biosynthesis. **Nucleic Acids Research**, v. 36, n. 6, p. 1813–1825, 2008.

SHIBUE, R. et al. Comprehensive reduction of amino acid set in a protein suggests the importance of prebiotic amino acids for stable proteins. **Scientific Reports**, v. 8, n. 1, p. 1–8, 2018.

SMITH, T. F. et al. The origin and evolution of the ribosome. **Biology Direct**, v. 3, n. 1, p. 16, 2008.

TAMURA, K. Origins and Early Evolution of the tRNA Molecule. **Life**, v. 5, n. 4, p. 1687–1699, 2015.

TAMURA, K. The Genetic Code: Francis Crick's Legacy and Beyond. **Life**, v. 6, n. 3, p. 36, 2016.

TRAIL, D.; WATSON, E. B.; TAILBY, N. D. The oxidation state of Hadean magmas and implications for early Earth's atmosphere. **Nature**, v. 480, n. 7375, p. 79–82, 2011.

VAN DER LINDEN, M. G.; DE FARIAS, S. T. Correlation between codon usage and thermostability. **Extremophiles**, v. 10, n. 5, p. 479–481, 2006.

VARMA, S. J. et al. products of the acetyl-CoA pathway. **Nature Ecology &**

Evolution, n. Umr 7006, 2018.

VISHWANATH, P. et al. Ribosomal protein-sequence block structure suggests complex prokaryotic evolution with implications for the origin of eukaryotes.

Molecular Phylogenetics and Evolution, v. 33, n. 3, p. 615–625, 2004.

WATSON, J. D. et al. **Biologia Molecular do GeneArtmed Editora**, 2015.

Disponível em: <<http://books.google.pt/books?id=nHjl5lzBkXcC>>

WOESE, C. R. On the evolution of the genetic code. **Proceedings of the National Academy of Sciences of the United States of America**, v. 54, n. 6, p. 1546–52, 1965.

WOESE, C. R. et al. The molecular basis for the genetic code. **Proceedings of the National Academy of Sciences of the United States of America**, v. 55, n. 4, p. 966–74, 1966.

WOESE, C. R.; FOX, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. **Proceedings of the National Academy of Sciences**, v. 74, n. 11, p. 5088–5090, 1977.

WOESE, C. R.; KANDLER, O.; WHEELIS, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. **Proceedings of the National Academy of Sciences**, v. 87, n. 12, p. 4576–4579, 1990.

WONG, J. et al. Coevolution Theory of the Genetic Code at Age Forty: Pathway to Translation and Synthetic Life. **Life**, v. 6, n. 1, p. 12, 2016.

WONG, J. T.-F. A Co-Evolution Theory of the Genetic Code. **Proceedings of the National Academy of Sciences**, v. 72, n. 5, p. 1909–1912, 1975.

WONG, J. T. F. Coevolution theory of genetic code at age thirty. **BioEssays**, v. 27, n. 4, p. 416–425, 2005.

YANG, X.; GAILLARD, F.; SCAILLET, B. A relatively reduced Hadean continental crust and implications for the early atmosphere and crustal rheology. **Earth and Planetary Science Letters**, v. 393, p. 210–219, 2014.

ZHU, C. T.; ZENG, X. B.; HUANG, W. D. Codon Usage Decreases the Error Minimization Within the Genetic Code. **Journal of Molecular Evolution**, v. 57, n. 5,

p. 533–537, 2003.